DOCUMENT RESUME

ED 189 195 TM 800 380

AUTHOR Vander Lugt, Garret A.
TITLE An Analysis of the Tenth Grade MEAP Reading Test for Racial Bias.
PUB DATE 28 Mar 80
NOTE 23p.; Paper presented at the annual meeting of the Michigan Educational Research Association (Dearborn, MI, March, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; Black Students; Difficulty Level; Goodness of Fit; Grade 10; High Schools; *Item Analysis; *Racial Bias; *Reading Tests; *Test Bias; Test Items; Test Results; White Students
IDENTIFIERS *Michigan Educational Assessment Program; *Rasch Model

ABSTRACT

Various operational definitions of test bias are discussed including those which involve rater judgment, statistical methods which use an external criterion, and statistical methods in which no external criterion is used. It is concluded that the Rasch model provides a theoretically acceptable method for detecting item bias. The Rasch Model is used to analyze tenth-grade reading test results on the Michigan Educational Assessment Program. Test bias was suspected because White students mastered 14 of the 15 objectives while Black students mastered only 10. Rasch analysis showed no statistically significant bias for the entire test, although one objective (locating items in a category) was more difficult for Blacks than for Whites at the same achievement level, and another (using graphs) was more difficult for Whites. When both objectives were removed the effects were cancelled. It is concluded that it may be more productive to accept test results and concentrate instead on eliminating the inequities which influence the achievement disparity. (Author/CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED189195

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM | CS
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view. | UD

AN ANALYSIS OF THE TENTH GRADE MEAP READING TEST FOR RACIAL BIAS

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Presented by

Dr. Garret A. Vander Lugt
Director of Data Services
Kalamazoo Public Schools

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G. Vander Lugt
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Michigan Educational Research Association
Dearborn, Michigan
March 28, 1980

TM 800380

ABSTRACT

Various operational definitions of test bias are discussed including those which involve rater judgment, statistical methods which use an external criterion, and statistical methods in which no external criterion is used. The merits of the various definitions are discussed, and it is concluded that the Rasch model provides a "theoretically acceptable" method for detecting item bias. The Rasch model is used to analyze tenth grade MEAP reading test results of Black and White students in the Kalamazoo Public Schools. Little evidence is found which indicates that the difference in reading achievement between Black and White students is due to racial bias of the MEAP test. It is concluded that, although analysis of test bias may lead to the construction of more equitable tests, it is more important to use test results in an equitable manner and to provide educational experiences which will overcome the effects of past discrimination.

JUN 1 3 1980

AN ANALYSIS OF THE TENTH GRADE MEAP READING TEST FOR RACIAL BIAS

On the 1979 Michigan Educational Assessment Program (MEAP) tenth grade reading test, 67.3% of the tenth graders in the Kalamazoo Public Schools demonstrated mastery of 75-100% of the objectives that were tested, compared to 61.0% for students statewide. However, an analysis of the test results by race indicated that 77% of the White students demonstrated mastery of 75-100% of the objectives, but only 39% of the Black students demonstrated mastery of 75-100% of the objectives (see Table 1). Stated differently, a typical White student had mastered 14 of the 15 objectives, while a typical Black student had mastered only 10 of the 15 objectives (see Table 2).

Results such as those described above, lead to serious questions regarding the possibility of racial bias of the MEAP tests. The question of test bias is particularly important to the Kalamazoo Public Schools because it is under a court order to conduct "a systematic and rigorous examination of the total testing program" in the school system. The objective of this examination must be to minimize racial bias in assessment practices which lead to student placement in low-level academic programs (Green and Cohen, 1979, p. 309). Thus, the court order and the author's interest in the Rasch model provided the motivation for conducting the study that is described in this paper.

Definitions of Test Bias

It has been pointed out that test bias can be defined in a number of ways, and the absence of a precise, operational definition of what is meant by test bias can lead to a great deal of confusion due to lack of clear communication (Flaugher, 1978, p. 671). For purposes of this paper, a test is considered to be free of racial bias if both Black and White students

TABLE 1

RESULTS OF TENTH GRADE MEAP READING TEST
FALL 1979
PERCENT OF PUPILS IN EACH CATEGORY OF ACHIEVEMENT

	PERCENT OF OBJECTIVES MASTERED			
RACE	0-24%	25-49%	50-74%	75-100%
WHITE N=622	4%	6%	14%	77%
BLACK N=207	14%	22%	26%	39%

TABLE 2

TENTH GRADE MEAP READING TEST
FALL 1979
AVERAGE PERCENT ATTAINMENT
BY RACE

RACE	AVERAGE % ATTAINMENT	MEDIAN OBJECTIVES MASTERED
WHITE N=622	84.2%	14
BLACK N=207	61.0%	10

interact with the items on a test in the same way. The nature of the interaction is determined by the Rasch model. Furthermore, each objective on the tenth grade MEAP reading test is considered to be one "item". An "item" is considered to be answered correctly if 4 or 5 of the 5 questions for an objective are answered correctly.

The definition of test bias that was described in the preceding paragraph is only one of many definitions that have been proposed and used. One approach that is commonly used is to have a panel of experts examine test items for bias. This approach is commonly used in the construction of standardized achievement tests (The Psychological Corporation, 1978), but it is not clear that this approach leads to a test which is equally valid for all racial or ethnic groups. Attempts to provide more sophisticated definitions of test bias have led to the development of a number of statistical techniques. These techniques may be divided into two categories - those which use an external criterion and those which use an internal criterion. Techniques which involve the use of an external criterion are generally inappropriate for use with achievement tests because there is usually no external criterion which has been shown to be unbiased (Burrill, 1975). Consequently, only techniques that do not involve the use of an external criterion will be discussed.

Many techniques for detecting biased test items have been proposed for cases in which there is no external criterion. Three will be mentioned in this paper. One method involves calculating normalized $\bar{p}$-values ("deltas") for each of the items and for both the minority and nonminority groups. A linear regression line is then calculated using these delta values from the two groups, and items with points more than $1\frac{1}{2}$ standard error of estimate from the regression line are tagged for potential bias. A second method involves calculating point-biserial coefficients of correlation between each item and total score for each person in the two racial groups. A statistically significant difference between

the point-biserial coefficients of correlation obtained for minority students and nonminority students on an item is evidence that the item may be biased. A third method examines differences in the response pattern to the choices on multiple choice items. A chi-square test is used to determine statistically significant differences in response patterns between two or more racial or ethnic groups.

All of the three methods discussed above have limitations. For example, the first two methods may indicate bias where none exists or vice versa. This limitation is inherent in the mathematical procedures that are used by the two methods. The third method does not indicate the direction of bias (i.e., for or against minority students), and it is not clear that differential response patterns result in a differential ordering of students by ability. Furthermore, the actual use of these methods can lead to some rather disconcerting results. When these methods were applied to the pool of 3873 items that were used in constructing the 1978 edition of the Metropolitan Achievement Tests, approximately 38% of the items were identified as biased by one of the methods, but less than 8% of the items were identified as biased by at least two of the methods, and only 0.5% of the items were identified as biased by all three methods. Finally, when the statistically biased items were submitted to an advisory panel, the vast majority were rated as <u>not biased</u>, and the advisory panel rated about the same proportion of statistically unbiased items as being biased (Beck and Sklar, 1978).

One definition of test bias that is sometimes used, often implicitly, is based upon the difference between the mean score for two or more racial groups. Using this definition, the tenth grade MEAP reading test is biased against Black students according to the data in Tables 1 and 2. However, in the case of achievement tests, acceptance of this definition makes it impossible to use test data to prove the existence of unequal educational opportunities, since any difference between the mean score of minority students and the mean score of nonminority students could be attributed to test bias. Given the

inequities that exist in society as a whole, such an argument hardly seems plausable (Flaugher, 1978, p. 673). For this reason, this writer does not accept the data in Tables 1 and 2 as proof that the tenth grade MEAP reading test is racially biased.

Defining Test Bias Using the Rasch Model

The Rasch model for item analysis is the simplest of a class of models that may be subsumed under "latent trait theory" (Baker, 1977). According to the Rasch model, the probability that a test item will be answered correctly depends only on the difficulty of the item and the ability of the person responding to the item. Other models add more item parameters, such as a discrimination parameter and/or a guessing parameter. The Rasch model assumes that all items have equal discrimination, and it ignores the problem of guessing. Assuming that a group of items is an equally valid measure of a single trait for all persons, calibration of item difficulties is independent of the ability distribution in the group of persons in the calibration sample. Similarly, once a group of items has been calibrated, any subset of items can be used to estimate a person's ability (Wright and Panchapakesan, 1969). The specific form of the item characteristic curve for the Rasch model is

$$P = \exp(b + d) / (1 + \exp(b + d)),$$

where P is the probability of a correct response, d is item difficulty and b is person ability. (Note that there is a different d for each item and a different b for each person.)

Wright, Mead, and Draba have shown how the Rasch model can be used to detect and correct test item bias, and it is claimed that the method that they describe is superior to other methods which have been proposed for detecting item bias, such as those described in the preceding section (Wright,

Mead, and Draba, 1976). The basic concept underlying the method is that persons should interact with items in the same way regardless of the racial group to which they belong. Specifically, item parameters should be the same regardless of the racial group that is used to calibrate them. Since the item parameters (difficulties) are obtained in a way that is independent of ability distribution in the calibration group, no external criterion is required.

Stated simply, racial bias is detected by examining shifts in item difficulty and discrimination when the items are calibrated using different racial groups. If students from different racial groups interact with the items in the same way, there should be no shifts in either item difficulty or item discrimination. Wright, Mead, and Draba show that the following holds for each item on a test

$$Y_v = a_o + a_1 (b_v - d) + e_v$$

where

Y_v is the approximate residual for person v in the minority sample

b_v is the ability of person v as determined by the nonminority sample calibration

d is the item difficulty as determined by the nonminority sample calibration

a_o is the shift in the item difficulty

a_1 is the shift in the item discrimination

and e_v is the error for person v. Thus, the difficulty shift and discrimination shift for each item can be determined using a linear regression model.

Results of the Rasch Analysis

The 1979 results of White students in the Kalamazoo Public Schools on the tenth grade MEAP reading test were subjected to a Rasch item analysis using an unconditional maximum liklihood procedure with a correction for statistical

bias (Wright and Douglas, 1977). Table 3 shows that four of the fifteen objectives did not fit the Rasch model, and these objectives were eliminated. Table 4 shows the item difficulties that were obtained for the remaining items. After eliminating students who mastered no objectives, students who mastered all the remaining objectives, and students who did not fit the model, 314 White students out of 622 remained in the calibration sample.

Table 5 shows the difficulty shifts and the discrimination shifts that were obtained by performing the linear regression analysis on the results that were obtained by Black students in the Kalamazoo Public Schools. After removing students who mastered no objectives and students who mastered all objectives, 160 Black students remained. Although the analysis indicated no statistically significant bias for the entire test, objectives 14 and 15 were studied in greater detail because they exhibited the largest shift in difficulty.

Table 6 shows the effect of removing item 14, item 15, or both, on the difference between the mean achievement level for White students and the mean achievement level for Black students. When item 14 is removed, the difference between the mean scores increases, which indicates that item 14 is more difficult for Blacks at a given achievement level than for Whites at the same achievement level. When item 15 is removed the difference between the mean decreases, which indicates that item 15 is easier for Blacks at a given achievement level than for Whites at the same achievement level. When both items are removed the effects cancel each other out. Table 7 shows that the differences in the percent of objectives accomplished for White and Black students follows the same pattern (although this would not necessarily happen in all cases). Table 8 provides a short description of each objective or the MEAP tenth grade reading test.

Finally, objectives 14 and 15, along with objective 4, which showed very little shift in item parameters, were rated by five educators. Table 9 summarizes the ratings. Notice that, for the most part, none of the items were rated as

TABLE 3

RASCH ANALYSIS OF TENTH GRADE MEAP READING TEST
FALL 1979
GOODNESS OF FIT TEST

OBJECTIVE	Z-VALUE	REJECTED
1	1.221	No
2	3.828	Yes
3	1.151	No
4	1.553	No
5	1.460	No
6	1.527	No
7	1.211	No
8	3.464	Yes
9	.865	No
10	.350	No
11	6.267	Yes
12	8.610	Yes
13	1.306	No
14	.923	No
15	.502	No

TABLE 4

RASCH ANALYSIS OF TENTH GRADE MEAP READING TEST
FALL 1979
OBJECTIVE DIFFICULTIES IN LOGITS

OBJECTIVE	DIFFICULTY	STANDARD ERROR OF DIFFICULTY
1	.556	.131
2	--	--
3	1.087	.125
4	-.067	.142
5	-.536	.155
6	-.749	.162
7	-.260	.147
8	--	--
9	.414	.133
10	-.775	.163
11	--	--
12	--	--
13	.317	.135
14	-.030	.142
15	.043	.140

TABLE 5

RASCH ANALYSIS OF TENTH GRADE MEAP READING TEST
FALL 1979
RACIAL BIAS TEST

OBJECTIVE	DIFFICULTY SHIFT	DISCRIMINATION INDEX (Centered at 0)
1	-.039	-.243
2	--	--
3	-.168	.305
4	.034	-.145
5	-.133	.142
6	.472	-.066
7	.277	.262
8	--	--
9	-.241	.180
10	-.311	.027
11	--	--
12	--	--
13	.024	-.144
14	.845	-.300
15	-.622	-.032

TABLE 6

TENTH GRADE MEAP READING TEST
FALL 1979
MEAN RASCH ACHIEVEMENT LEVELS

OBJECTIVES OMITTED	WHITE N=314	BLACK N=160	DIFFERENCE
14 & 15	1.008	0.152	0.86 $\pm$ 0.09
14	1.024	0.097	0.93 $\pm$ 0.08
15	1.042	0.237	0.81 $\pm$ 0.08
--	1.055	0.178	0.88 $\pm$ 0.08

TABLE 7

TENTH GRADE MEAP READING TEST
FALL 1979
AVERAGE PERCENT ATTAINMENT

OBJECTIVES OMITTED	WHITE N=314	BLACK N=160	DIFFERENCE
14 & 15	76.50%	58.75%	17.75%
14	76.75%	57.78%	18.97%
15	77.07%	60.56%	16.51%
--	77.43%	59.69%	17.74%

TABLE 8

TENTH GRADE MEAP READING TEST
FALL 1979
DESCRIPTION OF OBJECTIVES

OBJECTIVE NUMBER	DESCRIPTION
1	Determine meaning from context of a passage
2	Use roots and affixes to understand words
3	Read specialized words related to everyday functions
4	Tell major incidents that occur in a reading passage
5	Select the best sentence describing content of passage
6	Select passages that show causation
7	Select passages showing characterization
8	Choose best conclusion to reading selection
9	Tell purpose of reading selection
10	Identify literary types, subject matter, selections, etc.
11	Discuss purposes for reading various types of material
12	Alphabetize randomly chosen words
13	Relate general nature of the contents of sub-headings
14	Locate specified items in a categorical listing
15	Use graphs, charts, tables, maps, etc.

TABLE 9

TENTH GRADE MEAP READING TEST
FALL 1979
RACIAL BIAS ACCORDING TO RATER JUDGMENT

OBJECTIVE	RATER	NO RESPONSE	NONMINORITY		NOT BIASED	MINORITY	
			DEFINITELY BIASED	PROBABLY BIASED		PROBABLY BIASED	DEFINITELY BIASED
4	A		1	2	1		1
	B				5		
	C				5		
	D				4	1	
	E			1	2	2	
	F			(3)	2	(3)	
TOTAL:			1	3 (3)	19	3 (3)	1
14	A				5		
	B				5		
	C				5		
	D				5		
	E				4	1	
	F				5		
TOTAL:					29	1	
15	A	1			4		
	B			(3)	1	1 (2)	(1)
	C				5		
	D				5		
	E				4	1	
	F				5		
TOTAL:		1		(3)	24	2 (2)	(1)

highly biased. These ratings indicate that the raters were less sensitive to possible racial bias than the statistical test, and it is consistent with the results of the rating procedure that was followed in constructing the 1978 edition of the Metropolitan Achievement Tests.

Conclusion

The Rasch model for item analysis was used to analyze 1979 tenth grade MEAP reading test results of students in the Kalamazoo Public Schools for racial bias. The analysis provided little or no evidence that the MEAP tenth grade reading test is racially biased. This finding, along with the inconsistent results that are obtained using other methods for detecting racial bias, suggests that excessive concentration on possible bias of achievement tests may not be worthwhile. It may be more productive to accept the results of achievement testing as indicating a real disparity between the achievement levels of Black and White students. Efforts then should be concentrated on eliminating the inequitable educational opportunities which have led to the achievement disparity, and to design educational programs which overcome the effects of past and present discrimination.

References

Baker, F. B. Advances in Item Analysis. Review of Educational Research, 1977, 47, 151-178.

Beck, M. D. and Sklar, J. H. An Assessment of Item "Bias" in the 1978 Metropolitan Achievement Tests. Paper presented at the Eastern Educational Research Association Meeting, Williamsburg, Virginia, March 1978.

Burill, L. E. Statistical Evidence of Potential Bias in Items and Tests Assessing Current Educational Status. Paper presented at the Fourteenth Annual Southeastern Conference on Measurement in Education, December 1975.

Flaugher, R. L. The Many Definitions of Test Bias. American Psychologist, 1978, 33, 671-679.

Green, R. L. and Cohen, W. J. An Evaluation of the Results of the School Desegregation Order in Michelle Oliver, et. al., Plaintiffs, v. Kalamazoo Board of Education, et. al., Defendants. Report to the Honorable Noel P. Fox, Chief Judge of the United States District Court, Western District of Michigan, Southern Division, August 1979.

Wright, B. D. and Panchapakesan, N. A. Procedure for Sample-Free Item Analysis. Educational and Psychological Measurement, 1969, 1, 23-48.

Wright, B. D., Mead, M. and Draba, R. Detecting and Correcting Test Item Bias with a Logistic Response Model. Research Memorandum Number 22. Statistical Laboratory, Department of Education, The University of Chicago, October 1976.

Wright, B. D. and Douglas, G. A. Conditional versus Unconditional Procedures for Sample-Free Item Analysis. Educational and Psychological Measurement, 1977, 37, 573-586.

TABLE 1

RESULTS OF TENTH GRADE MEAP READING TEST

FALL 1979

PERCENT OF PUPILS IN EACH CATEGORY OF ACHIEVEMENT

	PERCENT OF OBJECTIVES MASTERED			
RACE	0-24%	25-49%	50-74%	75-100%
WHITE N=622	4%	6%	14%	77%
BLACK N=207	14%	22%	26%	39%

TABLE 2

TENTH GRADE MEAP READING TEST
FALL 1979
AVERAGE PERCENT ATTAINMENT
BY RACE

RACE	AVERAGE % ATTAINMENT	MEDIAN OBJECTIVES MASTERED
WHITE N=622	84.2%	14
BLACK N=207	61.0%	10

TABLE 3

RASCH ANALYSIS OF TENTH GRADE MEAP READING TEST

FALL 1979

GOODNESS OF FIT TEST

OBJECTIVE	Z-VALUE	REJECTED
1	1.221	NO
2	3.823	YES
3	1.151	NO
4	1.553	NO
5	1.460	NO
6	1.527	NO
7	1.211	NO
8	3.464	YES
9	.865	NO
10	.350	NO
11	6.267	YES
12	8.610	YES
13	1.306	NO
14	.923	NO
15	.502	NO

TABLE 4

RASCH ANALYSIS OF TENTH GRADE MEAP READING TEST
FALL 1979
OBJECTIVE DIFFICULTIES IN LOGITS

OBJECTIVE	DIFFICULTY	STANDARD ERROR OF DIFFICULTY
1	.556	.131
2	--	--
3	1.087	.125
4	-.067	.142
5	-.536	.155
6	-.749	.162
7	-.260	.147
8	--	--
9	.414	.133
10	-.775	.163
11	--	--
12	--	--
13	.317	.135
14	-.030	.142
15	.043	.140

TABLE 5

RASCH ANALYSIS OF TENTH GRADE MEAP READING TEST

FALL 1979

RACIAL BIAS TEST

OBJECTIVE	DIFFICULTY SHIFT	DISCRIMINATION INDEX (CENTERED AT 0)
1	-.039	-.243
2	--	--
3	-.168	.305
4	.034	-.145
5	-.133	.142
6	.472	-.066
7	.277	.262
8	--	--
9	-.241	.180
10	-.311	.027
11	--	--
12	--	--
13	.024	-.144
14	.845	-.300
15	-.622	-.032

TABLE 6

TENTH GRADE MEAP READING TEST

FALL 1979

MEAN RASCH ACHIEVEMENT LEVELS

OBJECTIVES OMITTED	WHITE N=314	BLACK N=160	DIFFERENCE
14 & 15	1.008	0.152	0.86 ± 0.09
14	1.024	0.097	0.93 ± 0.08
15	1.042	0.237	0.81 ± 0.08
-	1.055	0.178	0.88 ± 0.08

TABLE 7

TENTH GRADE MEAP READING TEST

FALL 1979

AVERAGE PERCENT ATTAINMENT

OBJECTIVES OMITTED	WHITE N=314	BLACK N=160	DIFFERENCE
14 & 15	76.50%	58.75%	17.75%
14	76.75%	57.78%	18.97%
15	77.07%	60.56%	16.51%
--	77.43%	59.69%	17.74%